Civic Capacities Alignment: Theoretical Statement Introduced in U.S. Civic AI Field Context

- Alignment Traditions and the Missing Civic Layer -

(From the Preface of *Non-Agentic Civic AI – Illuminating America's Civic Digital Field*)

Christopher Hunt Robertson, M.Ed.

With the Collaborative Support of Advanced AI – ChatGPT, Perplexity, and Gemini/Google

Democracies cannot govern what they cannot see.

Over the past decade, AI alignment has become a wide-ranging and urgent conversation. Some thinkers focus on hypothetical future superintelligences and existential risk. Others document present harms in welfare systems, labor markets, and digital platforms. Still others examine how attention, autonomy, and democratic norms erode under commercial and technological pressure.

Each of these traditions illuminates something essential, and none should be ignored.

This book series' contribution, ***Civic Capacities Alignment***, stands alongside those efforts, but it does not attempt to replicate them. Its focus is narrower and more concrete: the civic conditions under which any alignment work - technical, regulatory, or philosophical - can succeed in a democracy.

Theoretical Statement

Civic Capacities Alignment **recognizes existing civic capacities and their potential for complementarity while respecting the established roles of democratic and legal institutions and governmental processes. Specifically, Civic Capacities Alignment focuses on supporting the *Citizenry's Voluntary Distributed Civic Capacity*, *Continuous-Inquiry Interdisciplinary Expert Committees*, and *Non-Agentic Civic AI* (NACAI) as complementary civic capacities.**

The alignment sought is not an alignment of machine values with civic values, nor an attempt to model, predict, or represent collective democratic conscience. It is a constitutional division of labor in which voluntary citizen authority, accountable continuous inquiry, technological illumination, and lawful institutional action retain distinct roles. Citizens require greater shared visibility into complex civic conditions; interdisciplinary committees require accessible, continuously organized evidence; and NACAI requires human conscience and democratic judgment to give its observations legitimate public meaning. Together, these complementary civic capacities support civic problem-solving and civic data archiving.

Within this framework, NACAI serves as *Civic Visibility Public Infrastructure* - an auditable, non-agentic infrastructure that preserves public records, organizes and reveals relevant conditions and aggregate patterns, identifies system-level pressures, and helps institutions and communities recognize emerging conditions earlier. It must not score individuals, recommend policy outcomes, trigger enforcement, or exercise delegated governmental judgment. Judgment, political deliberation, and public responsibility remain with human citizens and institutions.

The theorist proposes a *U.S. Franklin-Dewey-Citizenry Civic Data Observatory within the nonpartisan, civilian Library of Congress* as a potential institutional home for coordinating NACAI's civic visibility function. Its role must be archival as well as illuminating: preserving, authenticating, and making intelligible the evidentiary record in a public environment where civic data is reproduced, compressed, recombined, and interpreted at scale. Without provenance-based authentication - the verification of civic data's origin and integrity - democratic judgment cannot remain reliably grounded in evidence.

Core principles of Civic Capacities Alignment, presented in *Non-Agentic Civic AI – Illuminating America's Civic Digital Field*, include No Moral Outsourcing, Digital Nature Integrity, Amoral AI Civic Advantage, and Moral Visibility.

Non-Agentic Civic AI may be highly capable without becoming unaccountable civic power or eroding democratic self-government.

(End of Statement)

This constitutional division of labor is Deweyan in spirit. John Dewey understood democracy not simply as a system for selecting officials, but as a form of associated life sustained by communication, inquiry, and the continuing capacity of a public to recognize and address the consequences of collective action. Democratic intelligence, in this sense, is not the intelligence of an expert system acting on behalf of passive citizens. It is the shared human practice of bringing consequences into view, testing interpretations against experience, hearing those affected, revising provisional understandings, and deciding openly what ought to be done.

Non-Agentic Civic AI can assist that practice only on its proper side: it may help make conditions more observable, records more preservable, patterns more comparable, and uncertainty more explicit. It may not stand in for the public inquiry through which people determine the meaning of those conditions, deliberate over competing obligations, and authorize lawful action. The objective is not automated collective intelligence, but better democratic conditions for human collective intelligence.

It begins from a simple claim: if free societies cannot see clearly what they are doing at the speeds and scales now shaped by AI, no technical solution can keep them aligned. Moral visibility - the capacity to perceive emerging civic harms, uncertainty, accountability pathways, and system-level drift in time to deliberate - is not a luxury. It is infrastructure.

What follows briefly sketches four influential alignment and AI-ethics traditions, what each makes visible, and the civic gap this work is designed to address.

Long-Termist Alignment: Nick Bostrom and Existential Risk

Nick Bostrom's *Superintelligence* crystallized a powerful concern: future machine intelligences may outstrip human control and pose existential danger. His work operates on long time horizons and at a planetary scale, emphasizing technical control problems, misaligned objectives, and the need for coordination around advanced systems.
This book does not contest those risks. It asks a prior civic question. Before humanity worries about steering hypothetical superintelligences, can existing institutions reliably see what today's systems are already doing? Before debating the goals of future agents, can democracies track hazards created by present systems operating at machine speed in schools, hospitals, courts, media systems, and elections? Where long-termist alignment looks far forward, this work focuses on a prerequisite layer: the visibility and responsibility conditions within the societies alignment is meant to protect. If those conditions collapse, even excellent technical alignment work may arrive too late, or land inside institutions too blind to govern it well.

In this sense, AI appears here not first as an existential adversary, but as an amplifier of already-present civic weaknesses: fragmented shared reality, diminishing democratic visibility, and institutions required to decide under conditions they can no longer adequately perceive.

Justice-First Analysis: Virginia Eubanks and Structural Harm

Virginia Eubanks's *Automating Inequality* documents how data systems and automation can deepen inequality in housing, welfare, and public services. Her work is empirically grounded and morally clear, showing how opaque systems can quietly punish poor and marginalized people. It remains a vital account of structural harm.

This work shares that concern for human dignity, but extends the analysis from diagnosis to civic design. It treats moral visibility as a democratic requirement: the ability of institutions and publics to see where harm is forming, who is bearing it, how patterns propagate across systems, and who retains responsibility for a response.

Where Eubanks shows how current systems can fail vulnerable communities, this work asks what kinds of non-agentic civic instruments might help repair that failure: instruments that monitor systems rather than individuals; surface hidden harms, uncertainty, and overlooked stakeholders without judging or scoring anyone; and reveal structural risk before it hardens into entrenched injustice.

Eubanks diagnoses social harm. This book proposes public-works responses: civic infrastructure designed to strengthen responsibility, public understanding, and democratic visibility without replacing human judgment.

Surveillance Capitalism Critique: Shoshana Zuboff and Autonomy

Shoshana Zuboff's *The Age of Surveillance Capitalism* offers a sweeping critique of how digital platforms extract behavioral data and commodify prediction. Her analysis portrays the fusion of capital and computation as a regime that threatens human autonomy by treating lived experience as raw material for prediction and influence.

This work shares Zuboff's concern that people are increasingly treated as inventory. It diverges in emphasis. Rather than centering economic extraction alone, it focuses on the epistemic and civic consequences: how citizens, officials, and institutions perceive risk in an environment where manipulation, distortion, and automated amplification reshape public life faster than democratic processes can respond.

Where Zuboff asks who owns behavioral data and the futures it enables, this work asks an adjacent question: how can democracies regain enough shared sight to govern responsibly in that environment? Its answer is a civic blueprint rather than a political-economy critique: a non-agentic, rights-preserving ensemble of interpretive tools that illuminates hazards without centralizing power, scoring citizens, or directing action.

Attention and Design Ethics: James Williams and Moral Focus

James Williams's *Stand Out of Our Light* examines the attention economy and argues that persuasive technologies erode conditions necessary for individual freedom. His work is centered on personal agency, the capacity to attend to what one finds meaningful.

This book shares Williams's concern for moral attention and inner freedom, but situates it within a Benjamin Franklin–inspired ethic of public works. Individual conscience remains the non-delegable source of moral judgment. Under contemporary conditions, private attention alone is not enough. The systems surrounding conscience now operate at speeds and scales that can overwhelm even the most disciplined individual.

Where Williams calls for more humane design at the level of products and interfaces, this work seeks to institutionalize clarity as a public good. It imagines civic-scale institutions able to say: "Here is what we saw. Here is what we do not know. Here is what we refused. Here is who owns the decision."

In this way, expressions of conscience are given something increasingly rare in digital life: a stable place to stand.

The Missing Civic Layer

Taken together, these traditions illuminate technical control problems, structural injustice, economic extraction, and the erosion of attention. Each addresses an essential facet of the alignment challenge. Yet across their differences, one operating layer is rarely treated as a primary design target: the civic infrastructure through which democratic societies perceive, interpret, contest, and respond to emerging harms.

None of these approaches is principally designed to answer questions such as:

- How can public institutions remain functional when hazardous digital patterns form faster than human perception?

- How can dispersed consequences become matters that citizens and communities can recognize, investigate, and deliberate about together?

- How can democracies restore shared visibility without automating moral judgment or centralizing coercive authority?

- How might non-agentic tools strengthen human judgment, preserve the right to refuse, and make responsibility visible rather than diffuse it?

- How can a society develop greater civic visibility without making citizens more legible to power?

That final question requires particular care. In *Seeing Like a State*, James C. Scott warned that institutions may simplify complex social realities in order to make populations administratively legible, discounting local knowledge and enabling overconfident intervention from above. His warning is not an argument against public knowledge. It is a warning against confusing an administrative representation of society with the society itself, and against allowing simplified visibility to become an instrument of centralized control.

The purpose of the civic visibility proposed here is therefore not to make citizens more visible to power. It is to make consequential institutional patterns, public risks, uncertainty, and lines of responsibility more visible to citizens, communities, and accountable public servants. Its instruments must remain rights-bounded, publicly auditable, explicit about what they cannot know, and open to correction by people whose lived knowledge reveals what any formal system has missed.

In the language of this work, the missing layer is ***Civic Capacities Alignment,*** the constitutional principle that AI's civic capacities must remain aligned with the limited role democracy permits them to perform. Its practical expression is ***Civic Visibility Public Infrastructure*** **-** rights-bounded public institutions, procedures, data practices, and non-agentic instruments through which a society can develop and sustain moral visibility.

Moral visibility is the practical capacity of a free society to perceive consequential patterns, emerging harms, uncertainty, accountability pathways, and system-level drift in time to deliberate. It joins a Benjamin Franklin-inspired public-works ethic to a John Dewey-inspired democratic ethic of inquiry: institutions should enlarge the public's capacity to learn from shared consequences without transferring moral or political authority to technical systems. Franklin supplies the practical imperative to build durable civic instruments; Dewey supplies the democratic requirement that those instruments remain communicative, publicly intelligible, corrigible, and answerable to the people whose common life they illuminate.

This book sketches one possible expression of that infrastructure: a ***Franklin - Dewey - Citizenry Civic Data Observatory System***, built on two fixed commitments - human moral primacy and non-agentic design.

Its ambition is modest but durable. It seeks to extend perception rather than automate control; to make risks, uncertainty, and responsibility more legible without scoring citizens, directing conduct, or concentrating authority. Its governing division of labor is simple:

AI illuminates. Humans judge. Lawful institutions decide.

Like Franklin's public works, and consistent with Dewey's experimental democracy, such an architecture must be tested in the open: piloted in limited settings, publicly explained, independently audited, challenged by affected communities, revised in light of experience, and, where necessary, stopped.

This framework does not replace existing alignment work. It addresses a condition under which technical, regulatory, and philosophical alignment efforts can become democratically usable: societies must retain enough shared visibility to govern themselves.

Non-Agentic Civic AI is offered to fellow alignment thinkers not as a competing school, but as a practical invitation to widen alignment's design target - from models and incentives alone to the civic infrastructure that surrounds them. It invites exploration of how rights-preserving, non-agentic instruments might extend early shared visibility across institutions during future public emergencies and other fast-moving civic crises, without automating judgment or centralizing power - giving humans a clearer map of risk while leaving the hardest choices in their hands.

This work does not claim to solve alignment. It proposes a public-works framework designed for open testing by any alignment agenda that takes democratic self-government seriously.

Civic Visibility Public Infrastructure must not make citizens legible to power; it must make power, risk, uncertainty, and institutional responsibility legible to the citizenry.

People must be able to collectively see civic digital signal disturbances early enough to make timely, informed decisions that may lessen harmful societal impacts and mitigate costly damages.

In the United States, federal, state, and local governments already commit vast public resources to recovering from digital disruptions in the public sphere, costs borne by taxpayers and the communities they serve. Prevention cannot eliminate every danger. But earlier shared visibility can help democratic institutions recognize consequential patterns, investigate them openly, and act before preventable harms cascade.

As Benjamin Franklin might remind us, "An ounce of prevention is worth a pound of cure." Let us build the ***Civic Visibility Public Infrastructure*** now required in the Age of Humanity with AI - not to make citizens more legible to power, but to ensure that people can see power, risk, uncertainty, and institutional responsibility clearly enough, soon enough, to sustain self-government.